昆虫学报 *Acta Entomologica Sinica*，October 2006，49(5)：726－732 ISSN 0454-6296

一个新的家蚕转录相关锌带蛋白基因的克隆及序列分析

冯 姗[1]，张耀洲[2,*]

（1. 浙江大学生物化学研究所，杭州 310029；2. 浙江理工大学生物化学研究所，杭州 310018）

摘要：锌带蛋白（zinc ribbon protein）是锌指类蛋白的一种，它的 Cys4 Zn(2＋)结合位点由 3 个 β_2 片层折叠而成，而不是 α 螺旋结构。锌带结构与锌指结构同为转录因子结合核酸的结构域，锌带蛋白作为转录相关因子在调节基因表达活性等方面具有重要作用。在对家蚕 *Bombyx mori* 蛹 cDNA 文库测序中，发现一个新的编码家蚕锌带蛋白基因的 EST 序列（GenBank 登录号 DY230964），以此序列为信息探针检索家蚕 EST 数据库，通过同源筛选，获得一个新的家蚕锌带蛋白基因 cDNA 全序列并经 RT-PCR 检测和克隆、测序验证，结果表明与电子克隆序列完全一致。我们将其命名为 *BmZNRD*1（Zinc Ribbon Domain Containing 1）（GenBank 登录号 DQ432055）。该基因全长为 675 bp，由 363 bp 的开放阅读框序列（ORF）、10 bp 的 5′端非翻译区序列（5′ UTR）和 302 bp 的 3′端非编码区序列（3′ UTR）组成，其编码的 120 个氨基酸序列与其他真核生物间具有较高的同源性（达 60％左右），预测分子量为 13.54 kD，等电点为 6.8。*BmZNRD*1 编码的氨基酸序列是一种锌带蛋白，推测有 2 个功能结构域，分别是位于 N-端的 $Cx_2Cx_{14}Cx_2C$ 和 C-端的 $Cx_2Cx_{24}Cx_2C$，其中 C-端保守氨基酸序列 $Cx_2Cx_6Yx_3QxRSADEx_2TxFx_2Cx_2C$ 在生物进化中保守性很高，从酵母、果蝇、线虫到两栖类、哺乳类都有发现该结构域的存在，与酵母 RNA 聚合酶 A 亚单位 9 和转录相关蛋白有很高的相似性，推测其具有相同的功能。将 *BmZNRD*1 基因 cDNA 序列与家蚕基因组序列进行比对，结果表明该基因具有 3 个外显子，2 个内含子，外显子/内含子边界符合经典的 GT-AG 规则。

关键词：家蚕；锌带蛋白基因；电子克隆；基因克隆；序列分析

中图分类号：S881.2；Q78 **文献标识码**：A **文章编号**：0454-6296(2006)05-0726-07

Cloning and sequence analysis of a novel transcription-associated zinc ribbon protein gene from silkworm，*Bombyx mori*

FENG Shan[1]，ZHANG Yao-Zhou[2,*] （1. Institute of Biochemistry，Zhejiang University，Hangzhou 310029，China；2. Institute of Biochemistry，Zhejiang Sci-Tec University，Hangzhou 310018，China）

Abstract：The zinc ribbons are one group of diverse zinc finger proteins. Most zinc-ribbon domain is actually folded as three-stranded antiparallel β-sheets in their structure instead of finger-like helices. The zinc ribbon is a domain that functions as interaction modules binding to nucleic acids. The zinc ribbon proteins perform a broad range of functions such as replication and repair，transcription and translation as a transcriptional repressor. Through searching cDNA bank of silkworm pupae，a unique EST（GenBank accession no. DY230964）encoding the zinc ribbon protein was identified. In order to obtain full-length cDNA in *Bombyx mori*，we used this sequence as a probe and performed in silico cloning based on the *B. mori* EST database. A contig was thus assembled on the basis of several ESTs. It was also verified by RT-PCR. PCR product was purified and ligated into pGEM-T-easy vector. The recombinant clones were sequenced，which had the same sequence as the contig. We nominated this novel gene encoding zinc ribbon domain containing protein as *BmZNRD*1（GenBank accession no. DQ432055）. The full-length of the gene is 675 bp. It contains an ORF of 363 bp，10 bp nucleotide sequence in 5′ UTR（untranslated region）and 302 bp nucleotide sequence in 3′

基金项目：国家“863”项目（2005AA206120）；国家“973”项目（2005CB121006）

作者简介：冯姗，女，1982 年生，山东人，硕士研究生，研究方向为生物化学与分子生物学，Tel.：0571-86843190，E-mail：fs1003@hotmail.com

* **通讯作者** Author for correspondence，Tel.：0571-86843198，E-mail：yaozhou@chinagene.com

收稿日期 Received：2006-04-07；**接受日期** Accepted：2006-06-23

UTR（untranslated region）. It encodes 120 amino acids and is highly homologous to those of eukaryotic organisms previously reported（about 60%），with their predicted mass 13.54 kD and isoelectric point 6.8. The deduced amino acid sequence is one of zinc ribbon proteins with two putative functional domains，the N-terminal domain $Cx_2Cx_{14}Cx_2C$ and the C-terminal domain $Cx_2Cx_{24}Cx_2C$. The highly conserved amino acid sequence within the C-terminal domain is $Cx_2Cx_6Yx_3QxRSADEx_2TxFx_2Cx_2C$，which is well conserved through evolution，in organisms such as yeasts，*Drosophila*，nematodes，amphibians and mammals，and highly homologous to the yeast RNA polymerase A subunit 9 and transcription-associated proteins. An alignment of the cDNA sequence with the silkworm genome sequences revealed that there were 3 exons and 2 introns in the gene of *BmZNRD*1，and all of them conformed the canonical GT-AG rules.

Key words：*Bombyx mori*；zinc ribbon protein gene；in silico cloning；gene cloning；sequence analysis

发现和研究新基因是当前国际上研究的热点，使用生物信息学的方法是发现新基因和进行功能研究的重要手段。随着人类和多种模式生物基因组计划的实施，表达序列标签（expressed sequence tag，EST）在基因结构与功能研究中的应用日益广泛（Adams *et al*.，1991），各种生物信息学技术也应运而生并不断完善，充分利用生物信息学资源和技术研究基因组结构与功能、发现新的功能基因、揭示基因起源与分子进化规律等已成为后基因组时代的核心任务。基于EST的电子克隆（in silico cloning）策略，是近年来发展起来的一门快速克隆基因的新技术，其技术核心是利用计算机技术，依托现有的网络资源、EST数据库、核苷酸数据库、基因组数据库等，采用生物信息学方法（包括同源性检索聚类、序列拼接等）延伸EST序列，以期获得部分乃至全长cDNA序列（Gill and Sanseau，2000）。进一步利用RT-PCR的方法进行克隆分析、验证。随着EST数据库的进一步完善，电子克隆策略已成为克隆新基因的重要方法，并成功地应用于人类基因组的研究（O'Brien *et al*.，2000；张德礼等，2002）。与传统的基因克隆方法相比，具有快捷、成本低、针对性强等特点（Schaefer，1995；Feng *et al*.，2004），但也受到dbEST的EST数量和质量的限制（Prigent *et al*.，1998）。

转录调节蛋白在基因转录调控中具有重要作用，这类蛋白的一个重要特征就是它们与DNA的特异结合区常具有典型的氨基酸组成和结构，即基序（motif），其中一类重要的基序就是锌指结构（Miller *et al*.，1985）。锌指结构是通过疏水作用并结合Zn^{2+}从而稳定地折叠形成'手指'状的结构，含有该结构的蛋白质称为锌指蛋白（zinc finger protein，ZFP）（Evans and Hollenberg，1988）。已有研究表明，锌指蛋白是一类广泛存在于动植物及人类的DNA结合蛋白，大多数锌指蛋白是基因调控因子。它们不仅可结合于DNA、RNA（Liao *et al*.，1992）和DNA-RNA杂交体（Shi and Berg，1995），还能与其他锌指蛋白结合（Lee *et al*.，1993），控制生物体中基因的转录、翻译和蛋白质合成。迄今为止，已发现的锌指类蛋白达千余种，按照不同的分类标准有不同的分类方法。根据蛋白质的空间结构，可以分为C_2H_2型锌指（C_2H_2-like finger），高音谱号锌指（treble clef finger）和锌带（zinc ribbon）3种类型（Krishna *et al*.，2003）。目前虽然对锌指蛋白与DNA的作用有初步的了解，但是对两者之间的作用模式，特别是对锌指蛋白中氨基酸残基与DNA中核苷酸碱基特异性识别的化学码尚未确定，其主要原因是目前可供研究的锌指蛋白数目较少，得到的信息还远远不够丰富。因此表达、利用新的锌指蛋白开展研究，将丰富研究锌指蛋白的类型，为建立和完善锌指蛋白中氨基酸残基与DNA中核苷酸碱基特异性识别的化学码提供必要的信息。

我们利用在家蚕*Bombyx mori*蛹cDNA文库测序过程中发现的一个新的编码家蚕锌带蛋白基因的EST序列作为信息探针，BLAST检索GenBank的家蚕EST数据库，拼接同源的EST序列，获得一个新的家蚕锌带蛋白基因cDNA全序列并经RT-PCR验证。同时利用生物信息学的方法分析该cDNA序列的基因特征及所编码的氨基酸序列特征，为研究其功能奠定了基础。

1　材料与方法

1.1　材料、菌种和主要试剂

菁松×皓月家蚕蛹（化蛹后第3天）由浙江中奇生物药业股份有限公司提供。大肠杆菌TG1由本实验室保存，M-MLV cDNA Synthesis kit、Taq DNA聚合酶、限制性内切酶、T4 DNA连接酶、Trizol试剂、

RT-PCR 系统购自 Invitrogen 公司，PolyATtract kit、X-gal、IPTG、Ampicilline 为 Promega 公司产品。pHelix 载体为 Boehring Mannheim 公司产品。质粒提取 Kit 和产物回收 Kit 为北京鼎国生物技术有限公司产品。ABI Prism 3100-A 测序仪。其他化学试剂均为国产分析纯。

1.2 方法

1.2.1 家蚕蛹总 RNA 的提取与 mRNA 的纯化： 取冻存于 -80℃的新鲜蚕蛹，经清洗、75%乙醇消毒后，置研钵中加液氮迅速充分研磨，转移至离心管中，按照 Trizol 试剂盒说明书提取总 RNA。采用 PolyATtract kit 纯化 mRNA，操作步骤按 Promega 公司提供的说明书进行。

1.2.2 cDNA 文库的构建： 以 mRNA 为模板，利用 M-MLV cDNA Synthesis kit 合成 cDNA 的第 1 和第 2 链，将分离的双链 cDNA 片段在 T4 DNA 聚合酶作用下末端补平，插入 *Hin*dⅡ处理过的 pHelix 载体中。用 $CaCl_2$ 转化法导入到大肠杆菌 TG1 中（Sambrook and Fritsch，1989），涂布在含有 IPTG、X-gal 和氨苄青霉素（100 μg/ mL）的 LB 平板上，37℃培养过夜。随机挑选分隔良好的白色单菌落接种在 LB（Amp^+）液体培养基内，在良好通气的摇床上 37℃培养过夜，抽提质粒 DNA，经 *Eco*RⅠ和 *Hin*dⅢ双酶切，电泳鉴定，在 ABI Prism 3100-A 测序仪上完成测序。

1.2.3 基因序列的电子克隆与生物信息学分析： 在家蚕蛹 cDNA 文库的测序过程中，发现了一个新的编码家蚕锌指蛋白的 EST 序列（GenBank 登录号 DY230964）。根据该 EST 序列，采用生物信息学的方法，以 DY230964 为信息探针，以 BLAST 检索程序检索 GenBank 中的家蚕 EST 公共数据库，检出与起始序列同源性较高或有部分重叠的 EST 序列，把检出序列首尾拼接，组装为连续体（contig），以该连续体序列为被检序列再进行 BLAST 检索，反复拼接并校对，尽可能获得大片段的 contig 或全长 cDNA（Kent，2002）。分析候选 contig 的基因特征，如核苷酸序列结构、外显子、内含子以及潜在的启动子区域、转录因子结合位点等，推测编码的氨基酸序列，对蛋白质特性及功能结构域进行分析与预测。所用各种分析工具见表 1。

表 1 生物信息学分析的主要工具

Table 1 The main tools of analysis in bioinformatics

常用资源 Common resources	网址及软件 Web sites and softwares
开放性阅读框查找 ORF finder	http://www.ncbi.nlm.nih.gov/gorf/gorf.html
基本 BLAST 检索 BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
启动子查找 Promoter inspector	http://www.genomatix.de/products/promoterInspector.html
转录因子结合位点查找 Mat inspector	http://www.genomatix.de/
蛋白质结构域分析 Prosite	http://ca.expasy.org/
可能的启动子预测 NNPP	http://www.fruitfly.org/seq-tools/promoter.html

1.2.4 RT-PCR 扩增家蚕锌指蛋白基因片段，重组、克隆及序列测定： 根据电子延伸结果，设计上游引物 P1 和下游引物 P2，分别包含 *Bam*HⅠ酶切位点、起始密码子 ATG 和 *Hin*dⅢ酶切位点、终止密码子 TAA，上海生工生物工程技术服务有限公司合成。

P1：5′-CGGGATCCATGTCCACTGAAAGCC-3′；

P2：5′-GGGAAGCTTTTAGGAGTTTTCAGTTTC-3′

（下划线表示酶切位点）

以供试家蚕蛹 cDNA 为模板，用所设计的引物 P1 和 P2 扩增特异片段。PCR 反应程序：94℃预变性 5 min，94℃变性 1 min，57℃复性 1 min，72℃延伸 1 min，共 30 个循环，72℃延伸 10 min。反应完毕后取 5 μL 于 1.2%琼脂糖凝胶进行电泳分析。低熔点胶回收 PCR 产物并纯化，与 pGEM-T-easy 载体连接，转化大肠杆菌 TG1 后挑选阳性克隆，双酶切鉴定并测序。

2 结果

2.1 家蚕锌指蛋白基因的 cDNA 全序列及 RT-PCR 验证

通过对家蚕蛹 cDNA 文库测序，获得一个新的编码家蚕锌带蛋白的 EST 序列，以该序列为信息探针，筛选家蚕 EST 数据库并进行反复拼接和校对，拼接后的 cDNA 序列长度为 675 bp。用 ORF finder 对该 cDNA 的编码框进行预测，得到一个长 363 bp 的编码区序列，位于全序列的 11～373 位核苷酸，含有起始密码子 ATG 和终止密码子 TAA，是一个完整的编码框，共编码 120 个氨基酸，5′端非翻译区（UTR）为 10 bp，3′端非翻译区长 302 bp。用特异性引物进行 RT-PCR 验证，PCR 产物经琼脂糖凝胶电泳检测，出现约 380 bp 的单一条带，与预计大小相符（图 1）。用 pGEM-T-easy 载体克隆后测序，结果与

电子克隆结果一致。

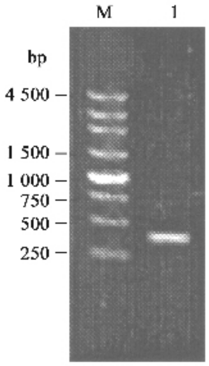


图 1　家蚕锌带蛋白基因 RT-PCR 扩增结果

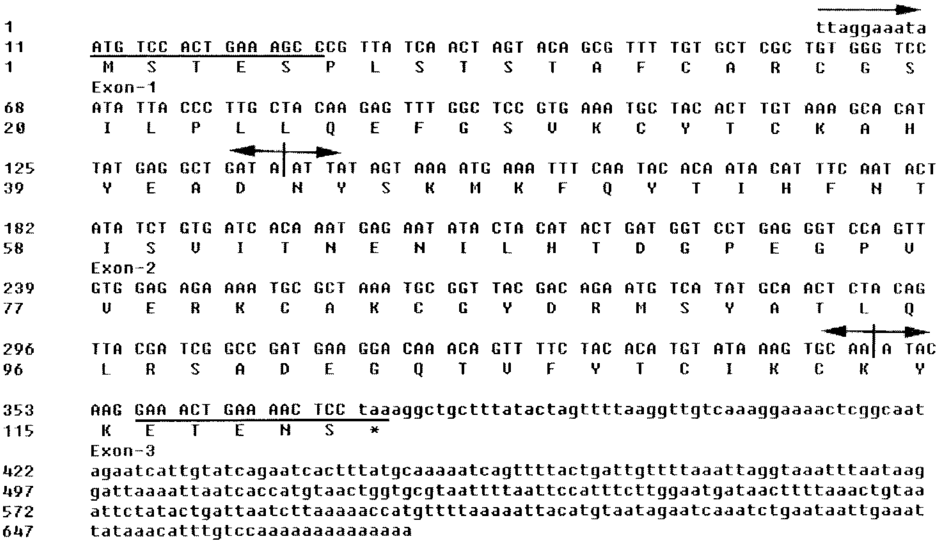
Fig. 1　The result of RT-PCR for zinc ribbon protein gene of *Bombyx mori*

M：DNA 梯度 TaKaRa 250 bp DNA ladder marker；1：RT-PCR 产物 RT-PCR product.

利用 NCBI 中 BLAST 检索软件进行检索分析，到目前为止，在家蚕中未发现任何具有同源性的序列，证明此 cDNA 是全新的。我们将其命名为 *BmZNRD*1（zinc ribbon domain containing 1）基因，GenBank 中登录号为 DQ432055。家蚕锌带蛋白 *BmZNRD*1 基因的 cDNA 全长序列及推测的氨基酸序列如图 2 所示。

2.2　家蚕 *BmZNRD*1 基因的结构分析

用生物信息学软件将 *BmZNRD*1 基因的 cDNA 全序列与家蚕基因组序列比对，结果显示：*BmZNRD*1 基因由 3 个外显子和 2 个内含子组成，翻译起始密码子 ATG 位于外显子 1 中，终止密码子 TAA 位于外显子 3 中(图 3)。其外显子与内含子连接区完全符合经典剪拼序列的 GT-AG 规则(表 2)。



```
1                                                                      ttaggaaata
11       ATG TCC ACT GAA AGC CCG TTA TCA ACT AGT ACA GCG TTT TGT GCT CGC TGT GGG TCC
1         M   S   T   E   S   P   L   S   T   S   T   A   F   C   A   R   C   G   S
         Exon-1
68       ATA TTA CCC TTG CTA CAA GAG TTT GGC TCC GTG AAA TGC TAC ACT TGT AAA GCA CAT
20        I   L   P   L   L   Q   E   F   G   S   V   K   C   Y   T   C   K   A   H

125      TAT GAG GCT GAT A|AT TAT AGT AAA ATG AAA TTT CAA TAC ACA ATA CAT TTC AAT ACT
39        Y   E   A   D    N   Y   S   K   M   K   F   Q   Y   T   I   H   F   N   T

182      ATA TCT GTG ATC ACA AAT GAG AAT ATA CTA CAT ACT GAT GGT CCT GAG GGT CCA GTT
58        I   S   V   I   T   N   E   N   I   L   H   T   D   G   P   E   G   P   V
         Exon-2
239      GTG GAG AGA AAA TGC GCT AAA TGC GGT TAC GAC AGA ATG TCA TAT GCA ACT CTA CAG
77        V   E   R   K   C   A   K   C   G   Y   D   R   M   S   Y   A   T   L   Q

296      TTA CGA TCG GCC GAT GAA GGA CAA ACA GTT TTC TAC ACA TGT ATA AAG TGC AA|A TAC
96        L   R   S   A   D   E   G   Q   T   V   F   Y   T   C   I   K   C   K    Y

353      AAG GAA ACT GAA AAC TCC taaaggctgctttatactagttttaaggttgtcaaaggaaaactcggcaat
115       K   E   T   E   N   S   *
         Exon-3
422      agaatcattgtatcagaatcactttatgcaaaaatcagttttactgattgttttaaattaggtaaatttaataag
497      gattaaaattaatcaccatgtaactggtgcgtaattttaattccatttcttggaatgataacttttaaactgtaa
572      attctatactgattaatcttaaaaaccatgttttaaaaattacatgtaatagaatcaaatctgaataattgaaat
647      tataaacatttgtccaaaaaaaaaaaaaa
```

图 2　家蚕 *BmZNRD*1 cDNA 的核苷酸序列和推测的氨基酸序列

Fig. 2　Nucleotide sequence and deduced amino acid sequence of the silkworm *BmZNRD*1 cDNA

大写字母为编码区；小写字母为非编码区；* 标记为终止密码子；预测的氨基酸序列以单字母标示在相应三联体密码下；箭头划分的区间表示外显子。The coding sequence is shown in uppercase letters, while the noncoding sequence is depicted in lowercase letters; the stop coden (taa) is indicated by asterisk (*); the deduced amino acid sequence is shown in single-letter code below the second base of each codon; the exons are divided by vertical lines and arrows at the boundary.

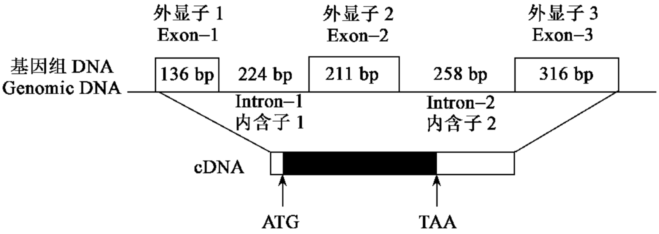


图 3　家蚕 *BmZNRD*1 基因的结构示意图

Fig. 3　Schematic representation of the silkworm *BmZNRD*1 gene

表 2　家蚕 *BmZNRD*1 基因的内含子与外显子结合区序列及内含子大小

Table 2　The junction sequences and sizes of *BmZNRD*1 gene exons and introns in the silkworm, *Bombyx mori*

内含子 Intron	**内含子大小** Intron sizes (bp)	5′**连接点** 5′ splicing site	3′**连接点** 3′ splicing site
1	224	TGATA $\boxed{gt}$ aag	tac $\boxed{ag}$ ATTAT
2	258	TGCAA $\boxed{gt}$ atg	ttc $\boxed{ag}$ ATACA

注 Notes：**大写与小写字母分别代表外显子与内含子序列；保守的二核苷酸剪切位点用方框表示。** Uppercase and lowercase letters indicate exon and intron sequences, respectively; the conserved splice donor and dinucleotide sequences are indicated in box.

2.3　家蚕 *BmZNRD*1 基因转录启动子序列的分析

用家蚕 *BmZNRD*1 基因的 5′端开放读框前 2 kb 基因组 DNA 序列在 NNPP 检索发现，存在 6 个可能性在 80%以上的启动子元件（表 3），但没有发现类似 TATA 盒的元件。此外，还存在 AP1 等转录因子识别和相结合的相关位点。而且在开放读框前存在 TAA 或 TGA 等多个终止子。

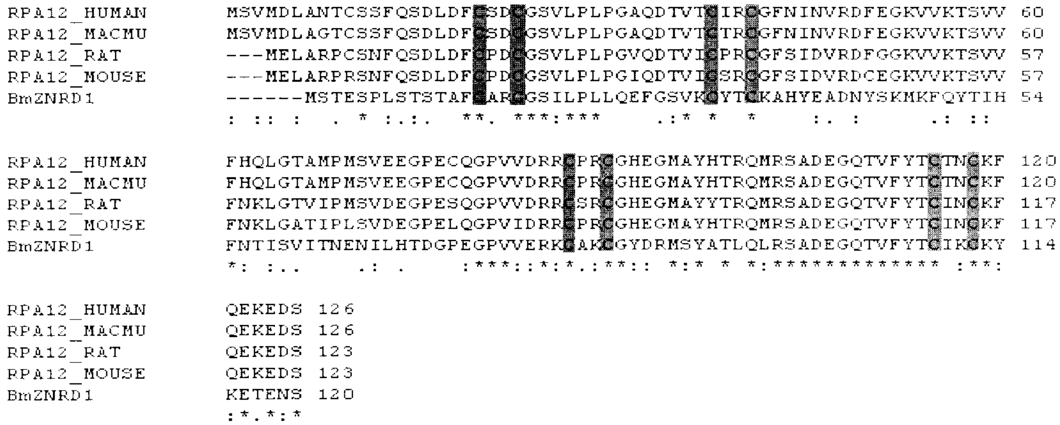
表 3　NNPP 预测的家蚕 *BmZNRD*1 基因可能的启动子

Table 3　Probable promoter sequences of *BmZNRD*1 gene predicted by NNPP web server

数目 No.	**起始位点** Start	**终止位点** End	**可能性** Score	**启动子序列** Promoter sequence
1	708	758	0.98	CGCGTTAGTATTAAAAAATCTCACAGATGGCGCTGTATTAAATAGTATGT
2	747	797	0.84	AAATAGTATGTATATTTCTGTCTCATTCTTTGCTGGCTTCATCCTACACA
3	932	982	0.90	GTTTTACGTTTTTAAAAGGTTCAAAGATAAAAATCAATTGATTGTTAATG
4	1 000	1 050	0.95	GTGTGGAATAAATAAATGCTTCCTATACTGCAGTTTGAGTTTAATCCGCA
5	1 082	1 132	0.85	CCTACTAATCTATATATGAGAAACAATGAAGATGACTGATTTTAGAGATC
6	1 192	1 242	0.93	GAGCAAGAGCTAAAAATGAGCTGTAGTAGACTGTAGTGGTTATATCTACA

2.4　家蚕 *BmZNRD*1 基因推测的氨基酸序列同源性分析

从 GenBank 等数据库中利用 BLAST 检索，分析与家蚕 *BmZNRD*1 基因预测的氨基酸序列同源性较高的相近氨基酸序列。发现与人 *Homo sapiens* RPA12 蛋白序列（Q9P1U0）有 60% 的相似性；与恒河猴 *Macaca mulatta*（Q5TM50）、家鼠 *Mus musculus*（Q791N7）、大鼠 *Rattus norvegicus*（Q6MFY5）中 RPA12 蛋白序列同源性分别为 60%、59% 和 61%。通过 Clustal W 进行多序列对比（图 4）可以看出，家蚕 *BmZNRD*1 基因的氨基酸序列 120 个氨基酸残基要比人、恒河猴、家鼠和大鼠的相应蛋白氨基酸序列短 3～6 个氨基酸残基，但其氨基酸序列的一致性在 40%左右，同源性达到 60%。并且决定锌指结构形成的半胱氨酸（C）残基序列与其他物种中相关蛋白的氨基酸序列保持高度一致，表明家蚕 *BmZNRD*1 基因的功能氨基酸残基（Cys）序列与其他物种的高度同源性。



```
RPA12_HUMAN     MSVMDLANTCSSFQSDLDFCSDCGSVLPLPGAQDTVTCIRCGFNINVRDFEGKVVKTSVV 60
RPA12_MACMU     MSVMDLAGTCSSFQSDLDFCSDCGSVLPLPGAQDTVTCTRCGFNINVRDFEGKVVKTSVV 60
RPA12_RAT       ---MELARPCSNFQSDLDFCPDCGSVLPLPGVQDTVICPRCGFSIDVRDFGGKVVKTSVV 57
RPA12_MOUSE     ---MELARPRSNFQSDLDFCPDCGSVLPLPGIQDTVICSRCGFSIDVRDCEGKVVKTSVV 57
BmZNRD1         ------MSTESPLSTSTAFCARCGSILPLLQEFGSVKCYTCKAHYEADNYSKMKFQYTIH 54
                : :: :  . * :.:.  **. ***:***    .:* *  *    :. :     .: ::

RPA12_HUMAN     FHQLGTAMPMSVEEGPECQGPVVDRRCPRCGHEGMAYHTRQMRSADEGQTVFYTCTNCKF 120
RPA12_MACMU     FHQLGTAMPMSVEEGPECQGPVVDRRCPRCGHEGMAYHTRQMRSADEGQTVFYTCTNCKF 120
RPA12_RAT       FNKLGTVIPMSVDEGPESQGPVVDRRCSRCGHEGMAYYTRQMRSADEGQTVFYTCINCKF 117
RPA12_MOUSE     FNKLGATIPLSVDEGPELQGPVIDRRCPRCGHEGMAYHTRQMRSADEGQTVFYTCINCKF 117
BmZNRD1         FNTISVITNENILHTDGPEGPVVERKCAKCGYDRMSYATLQLRSADEGQTVFYTCIKCKY 114
                *: :..    .: .    :***::*:*.:**:: *:* * *:************* :**:

RPA12_HUMAN     QEKEDS 126
RPA12_MACMU     QEKEDS 126
RPA12_RAT       QEKEDS 123
RPA12_MOUSE     QEKEDS 123
BmZNRD1         KETENS 120
                :*.*:*
```

图 4　由家蚕 *BmZNRD*1 基因推测的氨基酸序列与其他物种同源蛋白的比较

Fig. 4　Alignment of the deduced amino acid sequence of *BmZNRD*1 with homologous proteins from other organisms

RPA12-HUMAN：**人类** *Homo sapiens*, GenBank accession no. Q9P1U0；RPA12-MACMU：**恒河猴** *Macaca mulatta*, Q5TM50；RPA12-RAT：**大鼠** *Rattus norvegicus*, Q6MFY5；RPA12-MOUSE：**家鼠** *Mus musculus*, Q791N7；BmZNRD1：**家蚕** *Bombyx mori*, DQ432055. ＊：**相同的氨基酸** The identical amino acids；**黑色区域表示形成锌指结构的** 4 **个关键功能氨基酸**（Cys） Black area indicates the four key functional amino acids (Cys) in the zinc finger domain.

2.5　*BmZNRD*1 的蛋白质特性及功能结构域分析

家蚕 *BmZNRD*1 基因编码 120 个氨基酸，预测分子量为 13.54 kD，等电点为 6.8。利用 EMBL-EBI Interproscan 分析软件对所获得的氨基酸序列检索分析，发现家蚕 *BmZNRD*1 具有 2 个保守结构域，分别位于氨基酸残基序列的 13 ~ 35、77 ~ 117 氨基酸位置，其序列分别为 $Cx_2Cx_{14}Cx_2C$ 与 $Cx_2Cx_{24}Cx_2C$。N-端结构域由外显子 1 编码，在多种蛋白质中保守；C-端结构域为 ZF_TFⅡS_2（Zinc Finger TFⅡS_2-type），由外显子 2 编码，作为转录相关蛋白在多种生物体中高度保守，包括原核生物、酵母、线虫等。由图 4 可以看出，家蚕 *BmZNRD*1 的锌带结构与其他物种中相关蛋白质序列的功能域同源性非常高，而除此以外的其他区域同源性则比较低，推测其具有相似的功能。

3　讨论

我们应用在家蚕蛹 cDNA 文库测序过程中发现的一个 EST 片段作探针，利用日益增长的家蚕 EST 数据库以及电子克隆等生物信息学技术，通过将互为重叠的家蚕 EST 序列拼接，首次成功获得了家蚕新基因 *BmZNRD*1 全长 cDNA 序列。在此基础上，我们在 *BmZNRD*1 基因开放阅读框（ORF）内设计了特异引物，通过 RT-PCR 扩增了 *BmZNRD*1 基因片段，测序结果显示，其基因序列与电子克隆所获得的结果完全一致，由此证明了利用 EST 数据库克隆的 *BmZNRD*1 基因是正确和可靠的。

家蚕 *BmZNRD*1 是一个具有锌带结构的新基因，其 cDNA 由 675 bp 碱基序列组成，其中包括 363 bp 的 ORF，10 bp 的 5′UTR 和 302 bp 的 3′UTR，编码 120 个氨基酸残基。通过与家蚕基因组序列对比，表明该基因具有 3 个外显子，2 个内含子，外显子/内含子边界符合经典的 GT-AG 规则。对 *BmZNRD*1 基因 5′上游序列分析，不存在一般真核生物基因普遍具有的典型 TATA 盒启动子，但在其上游 2 kb 范围内有 6 个可能的启动子序列，还存在 AP1 等转录因子结合位点，推测该基因的转录起始不是由 TATA 盒调控，可能在家蚕细胞中表达量不是很高。

对家蚕 *BmZNRD*1 基因编码的 120 个氨基酸序列的同源性比较分析表明，*BmZNRD*1 是一种锌带结构转录相关因子，具有 2 个相对保守结构域，分别是位于 N-端的 $Cx_2Cx_{14}Cx_2C$ 和 C-端的 $Cx_2Cx_{24}Cx_2C$，与其他物种中相关蛋白质同源性约为 60%，而且其功能结构域的 4 个 Cys 氨基酸残基完全匹配。一般说来，如果一个蛋白质序列和较多不同种属的蛋白质序列具有较高的同源性（30% 以上），那么提示待分析的蛋白质序列可能是相应家族的成员（张成岗和贺福初，2002），因此家蚕 *BmZNRD*1 基因编码的蛋白质与其他物种的转录延伸因子在功能上应具有相似性，是一种转录相关蛋白质。进一步的蛋白结构分析显示，Cys4 Zn（2 +）结合位点是由 3 个 β_2 片层折叠组成，与转录延长因子（transcription elongation factor TFⅡS）结合 DNA 的结构域锌带结构同源，高度保守氨基酸序列 $Cx_2Cx_6Yx_3QxRSADEx_2TxFx_2Cx_2C$ 与酵母 RNA 聚合酶 A 亚单位 9 和转录相关蛋白高度相似。转录相关蛋白质可影响转录引起系列反应，同时还可以直接影响其他基因的转录和表达，如果它发生变异，其受调控的基因可能都要受到影响，进而引起系列反应，因此克隆具有重要生理功能的新基因 *BmZNRD*1 对于揭示家蚕基因调控机理，丰富锌带结构蛋白基因的研究内容，进一步开展锌带蛋白在家蚕中的生物学功能奠定了基础。关于家蚕 *BmZNRD*1 基因编码蛋白质的生物学功能正在进一步研究中。

参 考 文 献（References）

Adams MD，Kelley JM，Gocayne JD，1991. Complementary DNA sequencing：expressed sequence tags and human genome project. *Science*，252：1 651 – 1 656.

Evans RM，Hollenberg SM，1988. Zinc finger：Gilt by association. *Cell*，52：1 – 3.

Feng YJ，Zhang HM，Jiang MG，Lan XW，2004. In silico cloning of full length cDNA of *Cryphonectria parasitica* ubiquitin conjugated enzyme gene（CpUBC）. *China Journal of Bioinformatics*，2(2)：5 – 9.［冯友军，张会敏，姜明国，兰秀万，2004. 栗疫病菌泛素结合酶基因（CpUBC）全长 cDNA 的电子克隆. 生物信息学，2(2)：5 – 9］

Gill RW，Sanseau P，2000. Rapid in silico cloning of genes using expressed sequence tags（ESTs）. *Biotechnol. Annu. Rev.*，5：25 – 44.

Kent WJ，2002. Blat-Blast-like alignment tool. *Genome Res.*，12 (4)：656 – 664.

Krishna SS，Majumda I，Grishin NV，2003. Structural classification of zinc fingers. *Nucleic Acids Research*，31(2)：532 – 550.

Lee JS，Galvin KM，Shi Y，1993. Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *Proc. Natl. Acad. Sci. USA*，90：6 145 – 6 149.

Liao XB，Clemens KR，Tennant L，Wright PE，Gottesgeld JM，1992. Specific interaction of the first three zinc fingers of TFⅢA with the internal control region of the *Xenopus* 5S RNA gene. *J. Mol. Biol.*，223(20)：857 – 871.

Miller J，McLachlan AD，Klug A，1985. Repetitive zinc-binding domains in the protein transcription factor ⅢA from *Xenopus* oocytes. *EMBO*，4：

1 609 – 1 614.

O'Brien KP, Tapia-Páez L, Ståhle-Bäckdahl M, Kedra D, Dumanski JP, 2000. Characterization of five novel human genes in llq13-q22 region. *Biochem. Biophys. Res. Commune.*, 273(24): 90 – 94.

Prigent C, Gill R, Trower M, Sanseau P, 1998. In silico cloning of a new protein kinase, Aik2, related to *Drosophila aurora* using the new tool: EST Blast. *In Silico Biology*, 1(2): 123 – 128.

Sambrook J, Fritsch EF, 1989. Molecular Cloning: A Laboratory Manual. 2nd ed. New York: Cold Spring Harbor Laboratory Press.

Schaefer BC, 1995. Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. *Anal. Biochem.*, 227(2): 225 – 273.

Shi YG, Berg JM, 1995. Specific DNA-RNA hybrid binding by zinc finger proteins. *Science*, 268(14): 282 – 284.

Zhang CG, He FC, 2002. Bioinformatics Methods and Practice. Beijing: Science Press. 63 – 93. [**张成岗, 贺福初**, 2002. **生物信息学方法与实践. 北京: 科学出版社**. 63 – 93]

Zhang DL, Sun XJ, Ling LJ, Chen RS, Ma DL, 2002. Molecular cloning, characterization, chromosomal assignment, genomic organization and verification of SFRS12 (SRrp508), a novel member of human SR protein super family and a human homolog of rat SRrp86. *Acta Genetica Sinica*, 29(5): 377 – 383. [**张德礼, 孙晓静, 凌伦奖, 陈润生, 马大龙**, 2002. **人类** SR **蛋白超家族新成员——**SFRS12(SRrp508)**的基因克隆和特性分析. 遗传学报**, 29(5): 377 – 383.

(责任编辑: 黄玲巧)